*HOMOGENEOUS CATALOGS OF EARTHQUAKES**

By Leon Knopoff and J. K. Gardner

Institute of Geophysics, University of California (Los Angeles)

Communicated June 3, 1969

Abstract.—The usual bias in earthquake catalogs against shocks of small magnitudes can be removed by testing the randomness of the magnitudes of successive shocks. The southern California catalog, 1933–1967, is found to be unbiased in the sense of the test at magnitude 4 or above; the cutoff is improved to $M = 3$ for the subcatalog 1953–1967.

We consider the question of the completeness of a catalog of earthquakes for the purposes of statistical analysis. The catalog is usually a compilation of observations with a network of finitely spaced instruments, each with possibly different sensitivities. Small earthquake events which occur near a seismograph will be recorded while events of the same small magnitude occurring at distance from the nearest seismograph may not be recorded. Sufficiently large events will be tabulated no matter where they are located. Thus, there is a bias in any catalog against small shocks.

For any finite threshold of instrumental sensitivity and for any instrumental density, we assume there exists a threshold of seismic magnitudes, M_T, above which all events in a given region have a large probability of being recorded by at least one station of the network. Let the subcatalog with all events whose magnitudes $M \geq M_T$ be called a homogeneous catalog.

We create the subcatalog $C^{(n)}$ out of the original catalog C; $C^{(n)}$ is a chronological list of all events in C with magnitudes

$$M \geq M_n = M_L - n\Delta M,$$

where M_L is the largest earthquake magnitude in the catalog and ΔM is some selected interval of magnitudes. Evidently the chronologically consecutive events in $C^{(n)}$ need not be consecutive in $C^{(n+1)}$.

We take as the null hypothesis[1] the hypothesis that the sequence of magnitudes of chronologically successive events in a homogeneous catalog is random (R). We will take as the critical region for this hypothesis the condition $P = 0.01$ for type I errors.[2] Thus, the null hypothesis will be rejected for those catalogs $C^{(n)}$ with $M_n < M_T$. We call these catalogs $C^{(n)}(M_n > M_T)$ nonrandom (NR). NR catalogs are those whose probability of being generated by a random sequence of magnitudes is less than 0.01. We cannot say that the catalogs $C^{(n)}(M_n > M_T)$ are random. We can say, however, that these catalogs are not nonrandom (NNR). Thus NNR catalogs are those subcatalogs of the original from which all NR catalogs have been rejected. We find that all subcatalogs in which events whose magnitudes $M \geq M_n$ for all $M_n \geq M_T$ are NNR.

For each $C^{(n)}$, construct the matrix of transition numbers $T_{pq}^{(n)}$; $T_{pq}^{(n)}$ is the number of events in $C^{(n)}$ with magnitudes

$$M_L - (q + 1)\Delta M \leq M \leq M_L - q\Delta M,$$

which follow chronologically consecutively upon events of magnitudes

$$M_L - (p + 1)\Delta M \leq M \leq M_L - p\Delta M.$$

We may test the hypothesis by comparing the distribution $T_{pq}^{(n)}$ with a distribution of randomly selected magnitudes $\bar{T}_{pq}^{(n)}$. Tables of $\bar{T}_{pq}^{(n)}$ are called contingency tables.[2] Let

$$S_p^{(n)} = \sum_{q=0}^{n-1} T_{pq}^{(n)}$$

be the total number of events in $C^{(n)}$ with magnitudes

$$M_L - (p + 1)\Delta M \leq M \leq M_L - p\Delta M,$$

not counting the first event. It is not difficult to see that $S_p^{(n)}$ events distributed in proportion to the actual occurrence $S_q^{(n)}$ will lead to the distribution

$$\bar{T}_{pq}^{(n)} = \frac{S_p^{(n)}\, S_q^{(n)}}{R^{(n)}},$$

where

$$R^{(n)} = \sum_{p=0}^{n-1} S_p^{(n)}$$

is the total number of events in $C^{(n)}$ not counting the first event.

The comparison of $T^{(n)}$ with $\bar{T}^{(n)}$ may be made by a χ^2 test in the usual way. The $\bar{T}$ and T matrices are compacted to a rectangular matrix so that all $\bar{T}_{pq}^{(n)} \geq 5$, where the elements along the boundary of the compacted matrix are the sums of all the matrix elements in the same row or column with magnitudes greater (or less) than or equal to it. If the compacted matrix is a rectangular array of $P \times Q$ elements, the number of degrees of freedom is $d = (P - 1) \times (Q - 1)$. We maximize d in the compaction.

The above test was made on the Pasadena Catalog of Southern California Events[3] which lists 10,404 events occurring in the "Southern California Statistical Area"[3] during the 34-year period 1934–1967. Over time, the network had a varying density and quality of stations; $M_L = 7^3/_4$ corresponding to the Kern County earthquake (July 21, 1952). We take $\Delta M = {}^1/_2$, corresponding to the method of reporting magnitudes in 1934 but later refined. In Table 1, we give the probabilities of fit by $\bar{T}_{pq}^{(n)}$ of the various compacted observed sequences of magnitudes. The dramatic change in character of the table for catalogs $C^{(n)}$ truncated at magnitudes $3^3/_4$ and above, compared with those truncated below this value, is noteworthy. Since the magnitudes are reported at integer and half-integer values from 1934 to 1944, it is evident that $M = 3^3/_4$ is an artificial subdivision between the discrete values $M = 3^1/_2$ and $M = 4$. We infer that shocks with $M \leq 3^1/_2$ were not plentifully enough reported in the original catalog; distribution of these events through the 34-year chronology is significantly nonrandom at a probability level of less than 1 per cent. We conclude that the reporting of shocks with $M \geq 4$ has been adequately performed. The catalog

TABLE 1. *Analysis of the Southern California Catalog 1933–1967.*

n	M_n	$P \times Q$	d	$R^{(n)}$	Probability
5	$5^1/_4$	2×2	1	55	60%
6	$4^3/_4$	3×2	2	151	30
7	$4^1/_4$	3×3	4	503	30
8	$3^3/_4$	4×3	6	1409	1 –
9	$3^1/_4$	4×4	9	3454	$\ll 1$
10	$2^3/_4$	5×4	12	7328	$\ll 1$
11	$2^1/_4$	6×5	20	9525	$\ll 1$
12	$1^3/_4$	7×6	30	10,275	$\ll 1$
13	$1^1/_4$	8×6	35	10,383	$\ll 1$
14	$^3/_4$	8×6	35	10,398	$\ll 1$
15	$^1/_4$	8×6	35	10,398	$\ll 1$
16	$-^1/_4$	8×6	35	10,404	$\ll 1$

truncated to include shocks with $M \geq 4$ had only 1409 events remaining from the original list.

It is to be noted that the NNR hypothesis is assumed to apply both to shocks in the charging cycle[4] and to aftershocks. These two shock populations have different statistical distributions. In ordinary statistical operations, the two populations should be separated in some way, a problem that is burdened with pitfalls by virtue of the absence of a proper definition of an aftershock. Thus the success of the NNR hypothesis as applied to the entire catalog, and demonstrated in Table 1, indicates that we need not concern ourselves with the problem of the separation of the two problems at this stage. The applicability of the NNR hypothesis to both the aftershock and the charging cycle populations implies that the separation of the two can be made subsequent to this calculation, thereby reducing computational demands. In a later paper, the separation of the two populations will be made empirically, for a homogeneous catalog; for the separation algorithm that will be used, the concatenation of the two processes in this order is the less-consuming use of computer time.

In view of the foregoing, it may be possible to find some circumstances in which the NNR hypothesis does not apply to both populations equally as readily. One should look critically, perhaps, at catalogs containing large numbers of earthquake swarms. The Southern California Catalog does include some swarms in the Imperial Valley, in Walker Pass, and possibly elsewhere. If this contamination of the catalog were removed, the threshold of $M_T = 3^3/_4$ might be lowered somewhat; we guess that the amount of lowering will be small.

In view of the discussion about earthquake swarms, especially in the Imperial Valley, the possibility exists that the catalogs could be further subdivided into those corresponding to subregions of Southern California and into those corresponding to intervals of time shorter than the original 34-year interval. We have looked into the latter problem and have considered operations similar to those above, on the 1944–1967 and 1953–1967 catalogs. The date 1944 represents the time when the method of reporting magnitudes to the nearest 0.5 was changed to reporting magnitudes to the nearest 0.1. The date 1953 is significant in view of the improvement of instrumentation in southern California subsequent to the Kern County earthquake. The results are shown in Table 2. Evidently an improvement in M_T took place in 1953, but no improvement occurred

TABLE 2. *Probabilities that magnitudes in $C^{(n)}$ are generated randomly for three subcatalogs of Southern California Catalog.*

M_n	1934–1967 (%)	1944–1967 (%)	1953–1967 (%)
$5^1/_4$	60	70	
$4^3/_4$	30	20	
$4^1/_4$	30	70	30
$3^3/_4$	1−	1	30
$3^1/_4$	≪1	≪1	50
$2^3/_4$	≪1	≪1	1−
$2^1/_4$	≪1	≪1	≪1
$1^3/_4$	≪1	≪1	≪1
$1^1/_4$	≪1	≪1	≪1
$^3/_4$	≪1	≪1	≪1
$^1/_4$	≪1	≪1	≪1
$-^1/_4$	≪1	≪1	≪1

in 1944; in 1953, the threshold was lowered to about $M_T = 3$. There is an insufficient number of shocks $M \geq 5$ to construct an adequate χ^2 test on the contingency table for the 1953 catalog. Similar remarks may be made about the number of shocks with $M \geq 6$ in the two larger catalogs.

There is a rather remarkable agreement between the results presented in Table 2 and the qualitative impressions one may have about (1) the major changes in the seismographic network over the years and (2) the rough statistical evidence available about the quality of the tabulation of the catalog at various times. This agreement implies that the method presented above is indeed a technique for performing the qualitative evaluation quantitatively. These conclusions regarding M_T values are borne out by the descriptions of the changes in method of inclusion of small shocks in the catalog over time.[3]

* Publication no. 741, Institute of Geophysics, University of California (Los Angeles). This research was supported by U.S. Atomic Energy Commission contract AT(11-1)34-146.

[1] Cramer, H., *Mathematical Methods of Statistics* (Princeton University Press, 1946).

[2] Hoel, P. G., *Introduction to Mathematical Statistics* (New York: John Wiley, 1962), 3rd ed.

[3] Nordquist, J. M., *Bull. Seismol. Soc. Amer.*, **54**, 1003–1011 (1964).

[4] Burridge, R., and L. Knopoff, *Bull. Seismol. Soc. Amer.*, **57**, 341–347 (1967).